IZA DP No. 5976

The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment

Andries De Grip
Jan Sauermann

September 2011

The Effects of Training on Own
and Co-Worker Productivity:
Evidence from a Field Experiment

Andries De Grip
*ROA, Maastricht University
and IZA*

Jan Sauermann
ROA, Maastricht University

ABSTRACT

The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment[*]

This paper analyses the effects of work-related training on worker productivity. To identify the causal effects from training, we combine a field experiment that randomly assigns workers to treatment and control groups with panel data on individual worker performance before and after training. We find that participation in the training programme leads to a 10 percent increase in performance. Moreover, we provide experimental evidence for externalities from treated workers on their untreated teammates: An increase of 10 percentage points in the share of treated peers leads to a performance increase of 0.51 percent. We provide evidence that the estimated effects are causal and not the result of employee selection into and out of training. Furthermore, we find that the performance increase is not due to lower quality provided by the worker.

JEL Classification: J24, M53, C93

Keywords: training, field experiment, peer effects, productivity

Corresponding author:

Jan Sauermann
Research Centre for Education and the Labour Market (ROA)
Maastricht University
School of Business and Economics
P.O. Box 616
6200 MD Maastricht
The Netherlands
E-mail: j.sauermann@maastrichtuniversity.nl

1 Introduction

Estimating the returns of work-related training on productivity is a challenge for two reasons. First, the non-random selection of workers into training makes it difficult to identify the causal effect of training participation on individual productivity from other unobservable factors that drive participation in training as well as productivity. If this selection is not appropriately accounted for, the causal impact of training programmes on productivity can be overestimated (Bassanini et al. 2007). Second, it is difficult to find appropriate proxy variables for productivity. While wages are often used to estimate returns to training participation, the returns of employer-provided training on productivity should be higher than those on wages (Dearden et al. 2006); however, direct measures of productivity at the individual level are scarce.

Our contribution to the literature is twofold: We are the first to provide experimental evidence on (1) the causal effects of investments in training on productivity using individual performance information and (2) the externalities of training on co-worker productivity. A randomised field experiment enables us to measure the causal effects of investments in training on worker productivity. The field experiment was carried out in the call centre of a multinational telephone company in the Netherlands. We have detailed information on the contents, length, and purpose of the training programme, as well as unique panel data with administrative information on individual performance, to estimate the returns to training. Agents were randomly assigned to treatment and control groups. This exogenous variation in training participation is used to estimate the returns that are causally related to training and not to unobservable factors affecting both training participation and productivity. We find that agents perform 10 percent better after participating in the training programme than non-treated agents. We show that this effect is causal and not the result of employee selection into and out of training. Furthermore, we find that the performance increase is not due to a lower quality of work.

Workers belonging to the treatment group were trained over successive weeks. Moreover, due to the restricted capacity of the training centre the teams had to be split into two training groups. This time-varying treatment of the agents in a team allows for the further random assignment of agents into a first and a second training group. This enables us to identify possible externalities from training on untreated peers within a team. We find that a 10 percentage point increase in the share of treated agents leads to a performance increase of 0.51 percent for their untreated teammates.

From a policy point of view, an unbiased estimation of the impact of training on productivity is important for assessing the role of further training in the development of human capital. Previous studies have found rather mixed results, mostly depending on the method applied, as well as the measure of productivity used. We therefore categorise the research into three main strands. A large number of studies focus on the wage effects of work-related training and are extensively surveyed in Bartel (2000), Dearden et al. (2006), and Bassanini et al. (2007). After correcting for selectivity by applying fixed-effects estimation or instrumental variables regressions, these studies find returns of training ranging from zero (Goux and Maurin 2000) to 10 percent (Parent 2003). The disadvantage of using wages to measure the effects of training, however, is that wage increases after training only reflect the returns to the worker. The effect on productivity is thus likely to be underestimated under rent sharing.

The second strand of research uses large-scale surveys across firms, establishments, or industries. As a measure of firm productivity, most of these studies use the value added or sales of firms (Black and Lynch 2001; Dearden et al. 2006; Zwick 2006; Konings and Vanormelingen 2010) or direct measures of productivity within one sector of industry (Holzer et al. 1993). Both Dearden et al. (2006) and Konings and Vanormelingen (2010) show that the effects of training on firm productivity are about twice those on worker wages. When controlling for various sources of worker heterogeneity, the latter authors find that the productivity premium for a trained employee is, on average, 17 percent. Though large-scale surveys can have the advantage of providing a representative sample

for one or more sectors of industry, they inherently suffer from unobservable heterogeneity in the type as well as duration of training programmes and firm production processes (Ichniowski and Shaw 2009). Moreover, it is difficult, if not impossible, to find direct measures of productivity that are comparable across industries.

The third strand of research on the effects of training on productivity focuses on just a single firm or establishment, or comparable establishments within a sector of industry (Bartel 1995; Krueger and Rouse 1998; Liu and Batt 2007). Using direct performance measures and wages, they find mostly positive effects of training.

Overall, previous studies have estimated a range of treatment effects of training participation on productivity. At the same time, there is consensus that appropriate correction for selection into training matters for estimation. Although the use of field experiments has sharply increased over the last decade, only a few studies in the human capital literature exploit (quasi-) exogenous variations in training participation to estimate the effects of training. Leuven and Oosterbeek (2008) use information on the reasons for non-participation to artificially create treatment and control groups and find no significant effects from participation in training. The authors conclude that the usual methods to correct for selection into training do still overestimate the true effect of training. Similarly, Schwerdt et al. (2011) find no significant effects in their study analysing the effect of randomly assigned government-funded training vouchers.

This study exploits random assignment to training by means of a field experiment to overcome selectivity in training participation. By randomly assigning agents to treatment and control groups, we can estimate the average treatment effect (ATE) of training participation on individual productivity. We use panel data on individual worker productivity to estimate the effects of the training programme. To measure productivity, we use the key performance indicator used by the call centre to evaluate its call agents, that is, the average time needed to handle inbound customer calls. This performance measure has also been used in other studies in the call centre sector (Liu and Batt 2007; Breuer et al. 2010). We measure each worker's performance each week before and after

the training, which allows us to analyse short-run performance dynamics in the weeks after training. Using individual productivity as an outcome allows us to capture the total effect of training on productivity, while wage information captures only the share an agent receives.[1]

Despite a growing literature on peer effects in the workplace, the externalities of training have hardly been discussed in the literature on the impact of training on productivity.[2] The literature on human capital externalities originates from growth theory and has been applied to human capital spillover effects from workers' level of education (Moretti 2004). When the returns to training are estimated at the individual (worker) level without taking knowledge spillovers into account, returns will be underestimated. This also holds for a second type of externalities, peer effects due to social pressure (Falk and Ichino 2006). The identification of externalities, however, is empirically difficult. Previous studies used either quasi-random variations in group composition (Mas and Moretti 2009; De Grip et al. 2011) or the random assignment of subjects to groups (Guryan et al. 2009). While many field experiments have exploited changes in management behaviour induced by firm management (e.g., Shaw and Lazear 2008; Bandiera et al. 2007), we exploit the structure of the field experiment to estimate externalities. To achieve identification, teams were randomly split into two training groups. This random assignment of agents from the same team allows us to estimate the causal estimates of within-team externalities.

The training we analyse was a one-week course to train call agents in conversation techniques designed to decrease the average handling time of calls while maintaining call quality. Call centres have become a major sector of employment due to strong growth rates since the 1980s, facilitated by the increasing availability of information and com-

---

[1] A second advantage of using individual productivity instead of annual productivity data is that studies using annual data are more likely to underestimate the causal effects of training. If the yearly data do not contain information on the training period and use average productivity over the previous year, the returns to training are likely to be underestimated, even if selectivity is properly controlled for.

[2] To the best of our knowledge, the only exception is the study of Dearden et al. (2006) who aim to estimate externalities by comparing individual-level wage regressions with industry-level wage regressions that should capture externalities.

munication technology infrastructure. For the US, Batt et al. (2005) estimate that call centres employed about 4 million employees, roughly 3 percent of the total workforce. Although the current trend is to outsource call centres, most call centres are in-house (Batt et al. 2009). Work-related training is an important element of the call centre industry. In general, call agents receive hardly any initial vocational training, whereas the heavy use of information technology in in-house centres requires high investments in work-related training. From this perspective, we expect the returns on training in call centres will be high (Sieben et al. 2009).

The remainder of this paper is structured as follows: Section 2 provides an overview of the firm analysed in this paper and its workers and describes the experiment in detail. Section 3 presents our regression model and the estimation results. Additional evidence and tests on the robustness of our results are discussed in Section 4. Finally, Section 5 summarises the paper and presents concluding remarks.

2 Context of the experiment

2.1 Organisation of the department and worker tasks

The field experiment analysed in this paper was implemented in an in-house call centre of a multinational mobile network operator in the Netherlands. The call centre acts as a service centre for current and prospective customers. It has five departments segmented by customer group. To ensure a homogeneous production process, we focus on the largest department, that for private customers with fixed cell phone contracts.

The main task of call agents in this department is to answer customer phone calls. Customers contact customer service when they have problems, complaints, or questions. All agents take part in a training course when entering the department that enables them to handle basic types of calls. Throughout their careers, agents receive further training. These training programmes mainly focus on information in promotional campaigns,

communication, and information technology skills, as well as on handling more complex calls.

In the first period of our sample, 179 individuals were working in the department. Column (1) of Table 1 shows the descriptive statistics of all agents working in the department at the beginning of the period of observation. Most agents are part-timers. The average number of hours worked per week is 18.0, and only 22.4 percent of agents work 25 hours or more each week. A total of 29.1 percent of all agents are men, the average age of the agents is 32.6 years, and the agents have an average of 2.7 years of experience working for the firm.

– Table 1 about here –

The agents are organised into 10 teams. In general, all teams provide all services, that is, there is no team specialisation in handling certain types of calls or customers. All teams work on the same floor of the building. Work places are organised into work islands, with up to eight agents of a team sitting next to each other. Each team is led by a team leader who reports to the head of the department. The team leaders' main task is supervising and monitoring the agents of their team. They evaluate their agents regularly based on performance scorecards containing detailed information on key performance indicators. Agent pay is based on a single collective agreement. Agents are paid a base wage and receive an annual wage raise of zero to 8 percent, depending on an annual performance rating by the team leader. There are no other individual or team-related bonuses.

2.2 Training purpose, contents, and organisation

The training programme analysed in this study was intended for all agents of the department with some experience on the job. Its aim was to increase the efficiency of agents answering customer calls. Agents who were recently recruited were not selected for this

6

training (yet). The management had decided to organise the training to decrease the average time needed for handling calls, because the call centre had been performing below the targets set.

The training was organised as a week-long programme. Due to capacity constraints, only one group, with a maximum of 10 agents, could be trained at a time. The training took place in an in-house training centre located on a different floor in the same building. The programme consisted of 10 half-day training sessions that were held on five consecutive days, from Monday to Friday. In half of these sessions, agents were either formally trained by a coach or had group discussions assisted by a coach and their team leader. In these group discussions, agents discussed the skills they lacked, how their skills could be improved, and how agents could help each other on the work floor. In formal sessions, agents were trained in conversation techniques designed to decrease the average time needed for handling customer calls without lowering quality. This included, for instance, the way in which call agents gathered information from customers to resolve calls quickly.

The remaining sessions consisted of learning by doing, by either handling regular customer calls or listening to the calls of other call agents. During these sessions, incoming customer calls were routed to the training centre and agents handled these calls under the supervision of the training coach and their team leader.

2.3   The field experiment

In the economic literature on estimating the effects of training, the most evident problem is the potential correlation of unobservable factors with both training participation and the outcome variable. This study uses exogenous variations in training participation to identify the causal effect of training on individual productivity.

As shown in Figure 1, the field experiment consists of three periods. We observe 32 weeks, from week 45/2008 through week 24/2009. At the beginning of the first period

(pre-experiment), which lasted 17 weeks, agents were assigned to treatment and control groups. In this period, neither the treatment group agents nor the control group agents were trained. During the second, experiment period, the treatment group was trained consecutively over five weeks. After the experiment period (post-experiment, lasting 10 weeks), agents from the treatment group as well as their untreated peers from the control group worked as usual. Because the agents of the control groups were trained after the post-experiment period (from week 25/2009 onwards), we use data from week 45/2008 through week 24/2009 only. Agents who are part of the control group thus never participate in training throughout our observation period.

– Figure 1 about here –

The training had been designed for agents with some experience in the job. Out of the 179 individuals working in the department at the beginning of our observation period (week 50/2008), 86 were selected for the training.[3] Management focused the training programme on the more tenured agents for two reasons: First, agents should be experienced to benefit from this type of training and, second, management was aiming to lower the risk of losing its training investments through turnover.

Column (4) of Table 1 shows the differences between assigned and non-assigned agents. Assigned agents are an average of six years older, have longer tenure by more than three years, and perform better. While this selection is clearly non-random, it does not violate the assumption that assignment to the treatment group is exogenous. This is because assignment to the treatment and control groups is exogenous conditional on being assigned to the training programme. An advantage of the focus on more experienced agents is that the calls they handle are relatively homogeneous: Agents with longer tenure deal with all types of calls, whereas starting agents do not work on all types of calls yet.

---

[3]Agents not selected initially were assigned to training weeks after the training of the treatment group. Because these agents could not be randomly assigned to treatment and control groups, we do not consider them to be part of the control group. The effects we find therefore only hold for more experienced agents. As Sieben et al. (2009) show, training courses for more experienced agents are highly relevant in the call centre sector.

Conditional on being selected for the training programme, 37 of the 86 agents were randomly selected for participation in the treatment group. The remaining 49 agents, who were assigned to the control group, were trained after the post-experiment period, that is, after the end of the observation period. The differences in observable characteristics between the assigned treatment and control groups are relatively small, with none significantly different from zero (Column (7) of Table 1). For the estimation, however, agents were excluded if the management reassigned them from the control group to the treatment group, or vice versa, because they could not participate in the planned training week due to vacation or illness. Three agents from the treatment group and nine agents from the control group are thus excluded from the estimation sample. Column (10) of Table 1 shows that there is no statistical difference in observable variables between the treatment and control groups of our estimation sample.[4] The resulting estimation sample is an unbalanced panel of 74 agents during 32 weeks. In total, we have 1,859 agent-week observations on performance. On average, agents work 28 weeks out of the 32 weeks. The time-variant treatment in our experiment implies that the treatment and control groups differ during each specific week, that is, agents belong to the control group until they are treated. For this reason, teams were not only randomly assigned to the treatment and control groups, but also randomly assigned to the various weeks during the treatment period.

Identification of externalities

Although agents carry out individual tasks, social interactions can arise, influencing the treatment effect.[5] Agents belonging to the same team sit next to each other. Furthermore, they can communicate with their co-workers during team meetings and breaks.

---

[4]In addition, all tests were conducted by using the non-parametric Wilcoxon-rank test, which is suited for comparing discrete measures between small groups (Wilcoxon 1945). The results do not differ from those of the two-sided t-test.

[5]The production technology, however, is not necessarily without externalities. Agents may follow a strategy of providing low quality to score well on the main performance criteria (average handling time). If low-quality calls lead to a higher rate of customers calling repeatedly, other agents can be affected if they have to take more care of those customers. In practice, however, agents cannot easily shirk by providing low quality with a low average handling time because the quality of agents is regularly monitored by their team leaders, who can listen in on the calls. Moreover, if there was any shirking by providing low quality, these agents should be similarly distributed to the treatment and control groups.

In a survey held among the agents, 80 percent stated that they exchange work-related information either during official team meetings or at their workplace. Therefore, the likelihood of externalities is higher within teams than across teams. We do not preclude externalities between teams; however, given the physical distance on the work floor, these externalities should be less than externalities within teams.

Our field experiment allows us to identify these externalities of the training on untreated peers. For this purpose, we exploit the fact that eight of the 10 teams had to be split into separate training groups to match the size restriction of training groups in the firm's training centre, and the agents in these teams were randomly assigned to the two training groups.

Divided teams can thus be distinguished in the first and second training groups. In between the training weeks of the two training groups in which a team had been split, one or two groups of a different team were trained. During these weeks, the team consisted of exogenously chosen agents who were treated and randomly chosen agents who were not yet treated. This further randomisation enables us to distinguish between three groups of agents: (1) those who were trained first (first training group), (2) their teammates who were trained later (second training group), and (3) the agents in the teams in the original control group who will not be trained in the period we analyse. We exploit the fact that the agents of one team were randomly assigned to training groups to identify within-team externalities. Table 2 shows the means of observable characteristics for treatment group agents in a team's first training group ($N=13$) versus the means for agents who are in a team's second training group ($N=21$).[6]

– *Table 2 about here* –

All agents were informed about the training at the moment of their initial assignment. Throughout the experiment, information about the randomisation of the order of

---

[6]Both the t-test as well as the Wilcoxon-rank test do not show significant differences between the groups. It should be noted, however, that the sample size for this comparison is rather small.

the teams was given only to the head of the training department and the department manager. No information was given to the team leaders or the agents about the randomisation or the evaluation of the training programme. There was no need to do this, since all employees know that the size of the training centre is restricted. For this reason the training department always makes a training schedule in collaboration with the department managers when they start a new training program. The training centre is occupied most of the time. During the post-experiment period, agents from other departments of the call centre were trained in the training centre.

2.4   Measuring productivity

We measure agent productivity by means of the call centre's key performance indicator: the average handling time, defined as the average time an agent needs to handle a customer call (Liu and Batt 2007, Breuer et al. 2010). The handling time includes the time needed to talk to the customer, as well as the time needed to log the call in the customer database. Information on the average handling time of each individual agent i is available for each working week t. This performance measure does not contain information on the quality of calls. Section 4.3 introduces additional indicators of call quality and discusses the quality-quantity relation.

The department's aim is to improve performance by decreasing average handling time. Throughout this study, we therefore use the inverse of the average handling time (aht_{it}) multiplied by 100: $y_{it} = \frac{1}{aht_{it}} \cdot 100$. Since lower values of aht_{it} are interpreted as higher performance, this transformation allows us to interpret improvements in y_{it} as increases in individual productivity. The average performance of all agents assigned to training is 0.3307, which relates to 5.6 minutes for an average call (see Column (1) of Table 1).

There is substantial heterogeneity in individual performance within and between agents. This suggests that not only individual-specific characteristics, but also other, department-specific effects such as technical problems affect the individual productivity

of all agents working in the department. Compared to individual heterogeneity in productivity, however, variation over time is less important: While period fixed effects explain only about 12 percent of the overall variation in individual performance, worker fixed effects alone explain 43 percent of the overall variation in performance.

It is essential that our measure of productivity is comparable within and between agents. Calls are randomly assigned to agents. Agents have no direct influence on the types of calls they receive or the types of customers put through to them, and, therefore have the same probability of exceptionally long or short calls. Before talking to agents, however, customers must first state the purpose of their call. Based on this information, calls are routed to agents who have sufficient knowledge to resolve the call. The assignments of agents to types of calls can be changed at any time by management. Agents are often reassigned *ad hoc* if the structure of customer calls changes.

This assignment of calls to agents does not violate the assumption that our performance measure is comparable across time and agents. First, because agents are exogenously assigned to the treatment and control groups, the skill distribution of agents and thus the types of calls should, on average, be the same in the treatment and control groups. Second, we compare an agent's performance before and after the training intervention. Because the training did not focus on resolving different types of customer requests, calls assigned to agents after the treatment do not systematically differ from those assigned prior to the training.

3 Empirical analysis

3.1 Causal effect of training on individual performance

We observe an agent i's performance y_i and training participation d_i. The observed outcome can thus be written as

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot y_i(0) \tag{1}$$

where $y_i(1)$ and $y_i(0)$ denote performance in the treated and untreated states, respectively. The randomised assignment of agents to the treatment and control groups ensures the independence of treatment status and potential outcomes $E[y_i(j)|d_i = 0] = E[y_i(j)|d_i = 1]$ for $j = 0, 1$. The ATE τ (Rosenbaum and Rubin 1983) is thus identified by $\tau = E[y_i(1) - y_i(0)]$. The ATE can be estimated by performing a linear regression of individual i's performance y_{it} in week t on a treatment dummy d_{it}, which is defined as being one in each after-training period, and zero otherwise:

$$y_{it} = \alpha_i + \tau d_{it} + \beta_1 t_t + \beta_2 X_{it} + \beta_3 X_t + u_{it} \tag{2}$$

where α_i are individual fixed effects to account for individual heterogeneity remaining despite the experimental design, and X_{it} are covariates that are assumed to be independent from the treatment status d_{it}, such as working hours in week t and the type of shift the agent is working.[7] The variable t_t is a linear time trend that controls for trends in aggregate performance affecting all agents, and u_{it} is an idiosyncratic error term. To control for aggregate effects on performance (X_t), we include the overall number of calls

---

[7]We do not use agent tenure as a covariate in our regressions for two reasons. First, because the agents in our sample are all relatively experienced, there is not sufficient variation in individual tenure to control for tenure *and* for common time trends. Second, previous research shows that experience effects matter only for new agents (De Grip et al. 2011).

divided by the total number of full-time equivalent agents. Throughout this analysis, standard errors are clustered at the agent level.[8]

3.2   Baseline results

We provide first evidence of the treatment effect in Figure 2, which shows the average performance of agents in the treatment group. The treatment week is denoted week 0, with positive (negative) values of the x axis showing the tth calendar week after (before) the training. Agent performance appears to be, on average, higher in the weeks after the training than before the training. When the performance of untreated agents is not controlled for, treated agents perform significantly better after the training. Given the random assignment of agents to the treatment and control groups, this can also be shown by comparing the mean performance of agents in the treatment group with that of agents in the control group. Table 3 shows that agents in the treatment group are significantly more productive after the training, while there are no significant pre-treatment differences between the two groups.

– Table 3 about here –

– Figure 2 about here –

Table 4 shows the results when estimating Equation (2). The treatment dummy is defined as being one in all weeks after an agent has been trained, and zero otherwise. While agents from the control group thus always have a treatment dummy equal to zero, the share of treated agents increases in time with the growing number of groups that have been trained. Column (1) of Table 4 shows that agent performance after participation in the training is 11.3 percent higher than before the training, controlling for untreated

---

[8]In panels increasing in T with fixed N, the appropriate assumptions about standard errors can be crucial to the significance of the results (Bertrand et al. 2004). The results here were re-estimated, allowing for serial correlation in the standard errors. This does not change the size or significance of the results.

agents' performance. When, in addition, individual heterogeneity is controlled for by including worker fixed effects, the effect increases slightly to 12.4 percent (Column 2). Figure 2, however, suggests that aggregate trends matter. When controlling for a linear time trend, the effect decreases by about 3.6 percent to an estimated effect of 8.8 percent (Column 3 of Table 4).[9] The time trend itself is significant, suggesting that there is an overall positive trend in performance during the sample period.

– *Table 4 about here* –

The estimates in Column (3) of Table 4 show that the number of working hours do not have a significant impact on the performance y_{it}. In order to control for possible differences in the nature and amount of calls during peak hours, we use a variable that contains the share of peak hours in the total number of hours agent i worked in week t. The estimates show that a 10 percentage-point increase in the share of peak hours leads to a decrease in performance of 3.4 percent. This suggests that the calls during peak hours are more time consuming than the calls at other times of the day.[10] In order to control for overall pressure in week t, we furthermore control for the effect of the overall number of incoming calls in week t, divided by the number of full-time equivalent agents. This variable, however, does not significantly affect the performance of agents.

The preceding analysis makes the standard assumption that participation in training leads to a persistent shift in performance by including a dummy for participation in training. Given the weekly performance information at hand, we can exploit dynamic patterns of the treatment effect. Figure 3 shows the estimated shape of the treatment effect in the weeks after the training. Panel (a) of Figure 3 is based on an estimation of Equation (2) where the treatment variable d_{it} has been replaced by a set of dummies for each post-training week ($\sum_{t'=1}^{T} d_{it'}$), where $t' = 1$ denotes the first week after the

---

[9]To allow for more flexibility in time trends common to all agents, we also estimated Equation (2) including linear and squared time trends. The estimated treatment effect $\hat{\tau}$ then decreases slightly. The difference between the two estimates is not significantly different from zero.

[10]We also estimated regressions with similar variables to control for the timing of working hours, such as dummies for different working hours. These estimates show very similar results.

training. The untreated state $d_{it} = 0$ serves as the reference. It shows that the increase in performance in the first week after the training is not significantly different from zero. In the weeks thereafter, agent performance is significantly higher than before the training. Soon after, however, the estimated treatment effect gradually decreases and becomes insignificant from about the eighth week after the training. This decrease suggests that the estimate $\hat{\tau}$ can only be interpreted as an average effect over the whole post-treatment period.

– *Figure 3 about here* –

While this result seems to suggest that the treatment effect declines over time, there may be other mechanisms at work. From a human capital perspective, it is less reasonable that human capital acquired during the training depreciates within a short number of weeks. Moreover, the declining treatment effect may be due to a decline in the agents' motivation. From this perspective, however, it is difficult to understand that workers are least productive in the first week after the training. When the treatment effect would be driven by motivation, there should be a high positive performance effect in the first week. It is more likely that the decreasing treatment effect is driven by training externalities. If (yet) untreated agents are affected by the training of their peers, either due to knowledge spillover or peer pressure, their performance will increase when their co-workers have been treated. In this case, the treatment effect measured should be highest in the weeks shortly after the training and then constantly diminish, because agents from the control group are not actually fully untreated but indirectly affected by the training as well. Panel (b) of Figure 3 takes account of possible training externalities by showing the estimated shape of the treatment effects in the weeks after the training for the agents in the first training groups only, with the agents who will not be trained throughout the observation period as the control group. This figure shows a similar pattern as Panel (a). The treatment effect, however, remains significant for a longer period. Both the treatment and control groups in this setting are not affected by within-team externalities suggesting that Panel

16

(a) exhibits contamination through externalities leading to a decreasing treatment effect over time.

3.3 The role of social interaction

The preceding subsection estimates the ATE from participation in the training programme in the way that is standard in the training literature. An unbiased estimate $\hat{\tau}$, however, requires the stable unit treatment value assumption, hereafter SUTVA (Angrist et al. 1996). The most important implication of the SUTVA is that there are no externalities from treated workers on untreated workers; that is, the control group's performance ($y(0)$) is not affected by the treatment. In settings where individuals potentially interact, the SUTVA is violated. In the presence of externalities on untreated agents, the observed outcome changes to

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot (1 + d_i^* \tau_s) \cdot y_i(0) \tag{3}$$

where d_i, $y_i(0)$, and $y_i(1)$ are defined as in Equation (1). In addition, we allow the observed performance in the untreated state to be affected by the treatment of workers in the same team. The indicator d_i^* is defined as being one if a worker is untreated but exposed to treated workers from the same team. If an agent is untreated ($d_i = 0$) and working in a team in which all agents are untreated ($d_i^* = 0$), the potential outcome from Equation (3) simplifies to that from Equation (1). The same holds for treated agents who participated in the training ($d_i = 1$ and $d_i^* = 0$). The indirect treatment effect τ_s scales the size of the externality. If we assume that externalities are non-negative and affect untrained agents only, Equation (3) implies that, unless $\tau_s = 0$, the estimated treatment effect ($\hat{\tau}$) is underestimated in Equation (2). If direct and indirect treatment effects are stable, the true treatment effect is underestimated by τ_s.

In practice, the identification of social effects is cumbersome due to endogeneity (Manski 1993). In our setting, however, we can exploit the fact that team agents were

randomly assigned to two separate training groups. In between the training of two groups belonging to the same team, treated and untreated agents of the same team worked together for one or two weeks. During these weeks, a randomly selected group of the team that was treated worked with the remainder of the team that was untreated. Untreated agents of these teams were thus exposed to treated peers. Assuming that there are no peer effects between teams, the random assignment enables us to identify within-team externalities. We can do this in both a direct and an indirect way. The first column of Table 5 shows the results of a direct test of externalities. Here, we test whether the share of treated peers affects the performance of agents who were not yet treated. We therefore use precise information about an individual agent's shift to calculate the share of treated agents among all the agents with whom the agent worked during that shift. The share of treated peers is calculated for peers from the same team only. The results show that a 10 percent increase in the share of treated peers leads to a performance increase of 0.51 percent. This effect, however, is only statistically significant at the 10 percent level of significance.[11]

– *Table 5 about here* –

The second training groups having already increased their productivity through the externality effect of the first group being trained, we can expect a smaller additional effect when the second training groups finally receive the training themselves. We therefore test whether $\hat{\tau}$ is lower for agents who were exposed to treated agents before their own training. Column (2) of Table 5 shows the estimated treatment effect for agents who comprise the first group of their team to be treated, compared to that for the control group of agents who were in teams in which none of the agents were treated throughout the observation period. The estimated effect is slightly higher (9.9 percent) than when

---

[11]The regression presented in Column (1) of Table 5 does not contain the variable *share peak hours*. This is because both variables *share of peers treated* and *share peak hours* are based on the same hourly information. This causes a high between-agent within week correlation of the two variables and makes it impossible to identify the effect of *share peak hours* in this regression. This is supported by the fact that agents have only limited possibilities to choose shifts, and that the assignment to the first and second training group is random (Table 2).

estimating Equation (2) for all agents (8.8 percent; see Column (3) of Table 4). This suggests that the causal effect from training is underestimated because some agents were exposed to treated agents in their teams. Accordingly, the same estimation for second-group agents results in a lower estimate. Column (3) of Table 5 shows that their treatment effect is 6.8 percent. The difference between the two estimates, however, is not significantly different at the 5 percent significance level. Despite this, the results shown in Table 5 strongly suggest the existence of peer effects from treated agents on untreated peers in their team.

3.4 Returns to training

Standard theory predicts that firms invest in human capital if the expected returns from training investments exceed costs. Personnel information with specific information on the length and costs of training, as well as information on performance, allows to estimate the returns to training (Liu and Batt 2007). For the training analysed in this study, the costs of training per agent are defined as

$$C = \frac{(n_g \cdot w_a + w_l + w_c) * h}{n_g} \tag{4}$$

where n_g is the size of the training group; w_a, w_l, and w_c are the hourly wages for the agents, team leader, and training coach, respectively; and h is the length of the training in hours.[12] Because our performance measure is the time an agent needs, on average, to handle a call, and there are no other tasks in which the agent is involved, benefits are defined as the percentage increase in performance times the weekly workload and the agent's wage. The benefit B_d can be calculated for each week after the training d:

$$B_d = (\hat{\tau}_d \cdot aht_{id} \cdot n_{c,id}) \cdot w_a \tag{5}$$

---

[12]The wage costs of the team leader and training coach who supervise the training are included as direct training costs.

where $n_{c,id}$ is the number of calls an agent completes in after-training week d and $\hat{\tau}_d$ is the estimated treatment effect for this week as shown in Panel (b) of Figure 3. The term B_d is thus the decrease in the time a trained agent needs to handle a certain workload compared to an untrained agent in week d, multiplied by his wage.

Given the decreasing treatment effect over time, we use the actual average performance and actual number of calls made to calculate the returns to training for the firm. The firm will *de facto* retain the full benefits of the performance increase because agents' wages are almost entirely related to their years of tenure. Figure 4 shows the estimated curve of cumulated returns $\sum_{d=1}^{D} B_d$ until week D. The solid line shows that, for first-group agents, the cumulated benefits do not exceed the costs of the training in our sample period. When considering within-team externalities, however, the training reaches its break-even point in the eighth week after the training.[13] Figure 4 shows that in week 11 after the training, the overall return to the training is 37.7 percent higher when including externalities in the calculation of the benefits.[14]

– Figure 4 about here –

4 Additional evidence

4.1 First-week effects

One finding of the preceding analysis of the dynamics of the treatment effect is that agents perform better only starting from the second week after the treatment (see Figure 3). We can rule out the argument that this is due to selectivity, since over 90 percent of the agents who participated in the training also worked the following week. Another, yet

---

[13]Because half of the training consisted of actual customer calls, we calculated the costs of the training using 50 percent of the agents' wage costs. Considering the full wage costs of agents roughly doubles the training's costs.

[14]Externalities were calculated based on the estimates of the treatment effect for the first- and the second-group agents ($\hat{\tau}_1$ and $\hat{\tau}_2$) from Columns (2) and (3) of Table 5. We use the difference between the treatment effects as a measure of externalities. The externality effect on second-group agents is thus calculated as $(\hat{\tau}_2 - \hat{\tau}_1) \cdot aht_{id} \cdot n_{c,id} \cdot w_a$.

untestable, explanation for the low effect is that the low performance in the first week after the training is due to passing knowledge on to teammates.

A testable explanation, however, is fatigue. Most of the agents work part-time, with an average of about 20 hours per week. The training programme is one full-time week, that is, 38 hours. When using additional data on previous training programmes, we find that after rather short training programmes (less than 15 hours of training in one week), the difference between pre- and post-training is about the same, irrespective of the call agent's usual working hours (see Column (1) of Table 6). This is different for long training programmes where agents with relatively low numbers of working hours perform worse in the week after the program than before. In contrast, agents with long working hours do not exhibit this decrease in performance in the first week following the training. This result suggests that fatigue is an explanation for the slight decrease in performance in the first week after the training.

– Table 6 about here –

4.2 Personnel turnover and training effects

As in other call centres, the call centre analysed in this study experiences a high agent turnover. In total, we assigned 86 agents to the treatment and control groups, of which 12 were reassigned by the management or dropped out (see Section 2.3). Table 1 shows that the treatment and control groups do not differ with respect to observable characteristics such as age, gender, and tenure. This holds for the initial assignment, as well as for agents who eventually participated in the training.

Ordinary least squares estimation of Equation (2) results in an unbiased estimator of τ only if the probability of belonging to either the treatment or control group does not depend on unobservable characteristics that are correlated with both the likelihood to participate in the training programme and our performance measure, average handling

time. This type of selectivity occurs when agents with these characteristics are more likely to drop out of the treatment group than from out of the control group.

To test this hypothesis, we limit the estimation sample to agents that eventually participated in the training programme and stayed in the department throughout the observation period. This decreases the sample size from 74 agents in our main estimation sample to 55 agents. Of the 34 agents of the treatment group, 26 (77 percent) were trained and did not leave the call centre. For the control group, this limits the number of agents from 40 to 29 (73 percent). The main reason for the higher attrition in the control group is that those agents followed the training several weeks later, which increases the probability of dropping out before actually participating in the training. Calculating the attrition rate for treatment group agents who were staying at least until the last week of our full sample which contains longer panel information, we find there is hardly any difference in attrition between the treatment and control groups. In addition, we do not find a systematic pattern in the exit dates that reveals that agents may have left the department due to the upcoming training.[15] Table 7 shows that for this sample the estimated treatment effect decreases to 7.7 percent, which is lower but not significantly different from the estimated effect for the main sample (see Column (3) Table of 4).

– Table 7 about here –

4.3 Effects on call quality

Throughout this study, we used a transformation of the average length of calls as a measure of performance. Average handling time provides a clear measure of performance that allows us to quantify the returns to training precisely. One can, however, argue that individual productivity is a function of two dimensions. Calls can differ not only by

---

[15]There are potentially two dates on which selective agents may be more likely to leave due to the training. First, several months ahead of the first training, agents were informed that the whole department was going to be trained. Second, each agent received her schedule four weeks ahead. At neither of the two moments did exit rates appear to be higher than usual.

quantity, but also by quality provided by the agent. In this case, 'actual' productivity $y'_{it} = f(y_{it}, y^q_{it})$ is a function of the quantity-performance measure y_{it} and a performance measure for quality y^q_{it}. If y'_{it} is a non-monotonic function in y_{it}, improvements in average handling time will not necessarily translate into higher productivity. Since call quality is much more difficult to monitor than work speed, agents may aim to improve their handling times by providing lower call quality. However, agents cannot easily shirk by delivering low quality with a low average handling time because their calls are regularly monitored by their team leaders.

To test whether agents substitute higher performance y_{it} by providing lower quality y^q_{it}, we employ an alternative indicator for which we have data for each worker and week: the share of repeat calls. This measure is defined as the share of customers an agent talked to who called the call centre again within seven days. The firm uses this measure as an indicator of call quality because low customer satisfaction with the first call is the main reason for repeat calls.[16] On average, 23.6 percent of the customers called back within seven days after the first phone call (see Column (1) of Table 1). Because low values of repeat calls (rc_{it}) indicate higher performance, we define call quality as $y^q_{it} = \frac{1}{rc_{it}*10}$, with an average of 0.454.

The first column of Table 8 shows that training participation does not significantly affect call quality. The argument that the shorter average handling time after the training may be associated with lower call quality is therefore not supported by the data.

– *Table 8 about here* –

Having information on the quality of calls allows us to create a composite measure of productivity that incorporates both quantity and quality dimensions. We define y'_{it} as $y'_{it} = y_{it} \cdot (1 - rc_{it})$. Higher shares of repeat calls rc_{it} can thus be interpreted as a penalty in performance, since each additional percentage point in the share of repeat calls

---

[16]Note that the type (and potentially also the length) of repeat calls can be different from that of initial calls. Repeat calls are, however, not linked to certain agents. This implies that every agent has a given probability of receiving repeat calls.

(lower quality) relates to a lower composite productivity y'_{it}. The second column of Table 8 shows that the estimated treatment effect on the composite measure of productivity is 9.85 percent and statistically significant at the 1 percent level.

In addition to the information on the share of repeat calls, we use data gathered in a customer survey. Individual calls made by call agents were randomly selected and evaluated by customers on a scale from 1 (very bad) to 10 (very good). Customers gave grades on three different dimensions: the 'knowledge of the agent' (*Grade 1*), whether the 'agent understood the question' (*Grade 2*), and whether the agent had a 'solution to the problem' (*Grade 3*). Because the week t in which the call was made and the corresponding agent i are known, we are able to match this information to whether the agent was treated yet or not. The unit of observation for this analysis is thus a single call, and not agent-week information. Columns (3) through (5) of Table 8 indicate show the estimation results. They indicate that the training had a positive effect on the agent's knowledge and understanding capabilities. However, according to Column (5), the training had no effect on the agent's ability to provide a solution to the problem. Despite the fact that the analysis using customer-survey data is based on a rather small sample, the results are in line with the previous results showing that quality is not substituted at the cost of average handling time.

4.4 Heterogeneous treatment effects

The preceding analysis did not consider interaction effects between worker characteristics and treatment. This section analyses whether an agent's tenure and number of working hours affect the size of the estimated treatment effect.

First, more experienced agents may exhibit a different effect from the treatment compared to less experienced agents. On average, more experienced workers perform better than less experienced workers. When a ceiling exists in the potential productivity of call centre agents, less experienced agents have higher potential gains from attending the training. In contrast, complementarity in human capital acquisition can lead to greater

effects from participation in training for more experienced agents. Column (1) of Table 9 shows the treatment effect for agents with a tenure below the median tenure in our sample. Compared to Column (2), which shows regression results for agents with a tenure above the median, the estimated treatment effect is slightly higher for experienced agents (by 1.7 percentage points). The point estimates, though, are not significantly different from each other.

– Table 9 about here –

Second, agents with more working hours may experience a different treatment effect compared to agents with fewer working hours. Column (3) of Table 9 shows that the interaction effect between the number of working hours and the treatment dummy is significantly negative. This implies that the training participation of agents with more working hours has a lower effect on their productivity than the training of agents who work fewer hours. This result can be explained by either the greater fatigue of agents with more working hours or the selection of individuals into contracts with lower working hours based on unobservable characteristics such as ability and motivation. Students, for instance, may have contracts with shorter average working hours.

5 Conclusions

This study analyses the effect of training participation on worker performance by means of a field experiment held at a telephone company call centre. Agents had to participate in a compulsory five-day training programme. We randomly assigned agents to training groups, thereby generating exogenous variation in training participation. Regression results show that the performance of agents was about 10 percent better after having followed the training. This return is much higher than the returns to training found in studies that attempt to correct for selectivity. This may be explained by the fact that the effects of training on productivity are much greater than the returns measured by workers'

wages (Dearden et al. 2006; Konings and Vanormelingen 2010). However, as mentioned in the introduction, we also expect that returns to training in the call centre sector will be relatively high because the sector lacks proper vocational education for their core jobs. Furthermore, we find that the performance increase is not due to lower quality provided by the worker as indicated by a higher rate of repeat calls. Instead, training participation has a positive effect on customer ratings on agents' knowledge and understanding of the customers' question.

Moreover, we show that agents who did not participate in the training also improved their performance through externalities. These indirect effects may arise due to either knowledge spillover from trained to untrained workers or social pressure. We find that increasing the share of treated peers by 10 percentage points leads to a 0.51 percent increase in performance. We show that the effects measured are indeed caused by the training and not other effects, such as selective labour turnover.

Our finding of externalities in the workplace has important implications for the estimation of the effects of work-related training on wages using individual data. Even if estimates properly control for selective training participation, the estimated treatment effect is likely to be an underestimation of the true treatment effect, because externalities of training participation increase the productivity of those who have not been trained. Since the tasks carried out by call agents follow an individual production technology without team production, these externalities are probably even higher in other sectors of industry in which team work is more important than in the call centres.

The finding of externalities has important implications for firms' training strategies. When the externalities of training participation increase the productivity of workers who have not been trained, it may be more efficient to design training and the assignment by internalising the treatment externalities. This does, however, involve important questions for further research. Are there nonlinearities in the externalities and what is the optimal share of workers being treated? How and to what extent do externalities propagate further to other workers?

References

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification and Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2007): "Incentives for Managers and Inequality Among Workers: Evidence From a Firm-Level Experiment," *Quarterly Journal of Economics*, 122(2), 729–773.

BARTEL, A. P. (1995): "Training, Wage Growth, and Job Performance: Evidence from a Company Database," *Journal of Labor Economics*, 13(3), 401–25.

BARTEL, A. P. (2000): "Measuring the Employer's Return on Investments in Training: Evidence from the Literature," *Industrial Relations*, 39(3), 502–524.

BASSANINI, A., A. BOOTH, G. BRUNELLO, M. DE PAOLA, AND E. LEUVEN (2007): "Workplace Training in Europe," in *Education and Training in Europe*, ed. by G. Brunello, P. Garibaldi, and E. Wasmer, chap. 8-13. Oxford University Press, Oxford.

BATT, R., V. DOELLGAST, AND H. KWON (2005): "Service Management and Employment Systems in U.S. and Indian Call Centers," in *Brookings Trade Forum 2005: Offshoring White-Collar Work—The Issues and Implications*, ed. by S. M. Collins, and L. Brainard, pp. 335–372. Brookings Institution Press, Washington, D.C.

BATT, R., D. HOLMAN, AND U. HOLTGREWE (2009): "The Globalization of Service Work: Comparative Institutional Perspectives on Call Centers," *Industrial and Labor Relations Review*, 62(4), 453 – 488.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 119(1), 249–275.

BLACK, S. E., AND L. M. LYNCH (2001): "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity," *Review of Economics and Statistics*, 83(3), 434–445.

BREUER, K., P. NIEKEN, AND D. SLIWKA (2010): "Social Ties and Subjective Performance Evaluations: An Empirical Investigation," IZA Discussion Paper 4913, Institute for the Study of Labor (IZA), Bonn.

DE GRIP, A., J. SAUERMANN, AND I. SIEBEN (2011): "Tenure, spill-over effects, and output," Maastricht University, mimeo.

DEARDEN, L., H. REED, AND J. VAN REENEN (2006): "The Impact of Training on Productivity and Wages: Evidence from British Panel Data," *Oxford Bulletin of Economics and Statistics*, 68(4), 397–421.

FALK, A., AND A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–58.

GOUX, D., AND E. MAURIN (2000): "Returns to firm-provided training: evidence from French worker-firm matched data1," *Labour Economics*, 7(1), 1–19.

GURYAN, J., K. KROFT, AND M. NOTOWIDIGDO (2009): "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 1(4), 34–68.

HOLZER, H. J., R. N. BLOCK, M. CHEATHAM, AND J. H. KNOTT (1993): "Are Training Subsidies for Firms Effective? The Michigan Experience," *Industrial and Labor Relations Review*, 46(4), 625–636.

ICHNIOWSKI, C., AND K. SHAW (2009): "Insider Econometrics: Empirical Studies of How Management Matters," NBER Working Papers 15618, National Bureau of Economic Research, Cambridge, MA.

KONINGS, J., AND S. VANORMELINGEN (2010): "The Impact of Training on Productivity and Wages: Firm Level Evidence," IZA Discussion Paper 4731, Institute for the Study of Labor (IZA), Bonn.

KRUEGER, A., AND C. ROUSE (1998): "The Effect of Workplace Education on Earnings, Turnover, and Job Performance," *Journal of Labor Economics*, 16(1), 61–94.

LEUVEN, E., AND H. OOSTERBEEK (2008): "An Alternative Approach to Estimate the Wage Returns to Private-Sector Training," *Journal of Applied Econometrics*, 23(4), 423–434.

LIU, X., AND R. BATT (2007): "The Economic Pay-Offs to Informal Training: Evidence from Routine Service Work," *Industrial and Labor Relations Review*, 61(1), 75–89.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99(1), 112–145.

MORETTI, E. (2004): "Estimating the Social Returns to Higher Education: Evidence from Longitudinal and Repeated Cross-Section Data," *Journal of Econometrics*, 121, 175–212.

PARENT, D. (2003): "Employer-supported training in Canada and its impact on mobility and wages," *Empirical Economics*, 28(3), 431–459.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), 41–55.

SCHWERDT, G., D. MESSER, L. WOESSMANN, AND S. WOLTER (2011): "Effects of Adult Education Vouchers on the Labor Market: Evidence from a Randomized Field Experiment," IZA Discussion Papers 5431, Institute for the Study of Labor (IZA), Bonn.

SHAW, K., AND E. P. LAZEAR (2008): "Tenure and Output," *Labour Economics*, 15(4), 705–724.

SIEBEN, I., A. DE GRIP, D. VAN JAARSVELD, AND O. SØRENSEN (2009): "Technology, Selection, and Training in Call Centers," *Industrial and Labor Relations Review*, 62(4), 553–572.

WILCOXON, F. (1945): "Individual Comparisons by Ranking Methods," *Biometrics*, 1, 80–83.

ZWICK, T. (2006): "The Impact of Training Intensity on Establishment Productivity," *Industrial Relations*, 45(1), 26–46.

A   Figures

Figure 1: Overview of the field experiment
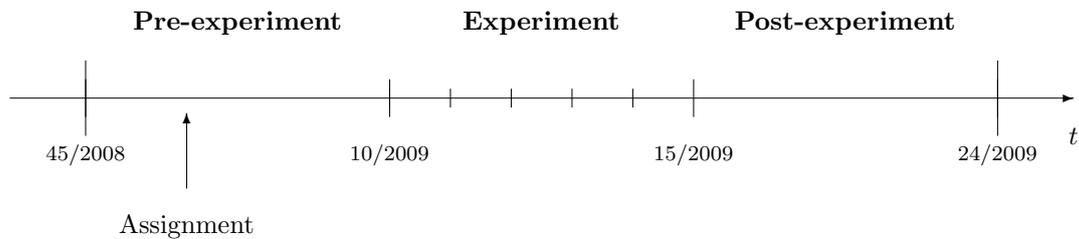
Note: This figure shows the design of the field experiment. The observations start in week 45/2008 and end in week 24/2009. The assignment to treatment and control groups took place in week 50/2008. Training of the treatment group was held in the experiment period.

Figure 2: Average performance of the treatment group before and after the training
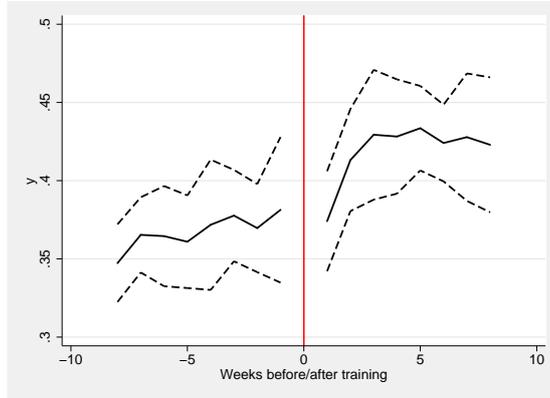
Note: This figure shows the average performance of agents in the treatment group in the weeks before and after the training week (solid line). The dashed lines show the 95 percent confidence interval. The training week is denoted as week 0.

Figure 3: Development of the treatment effect on performance over time

(a) Average effect (b) First training groups only
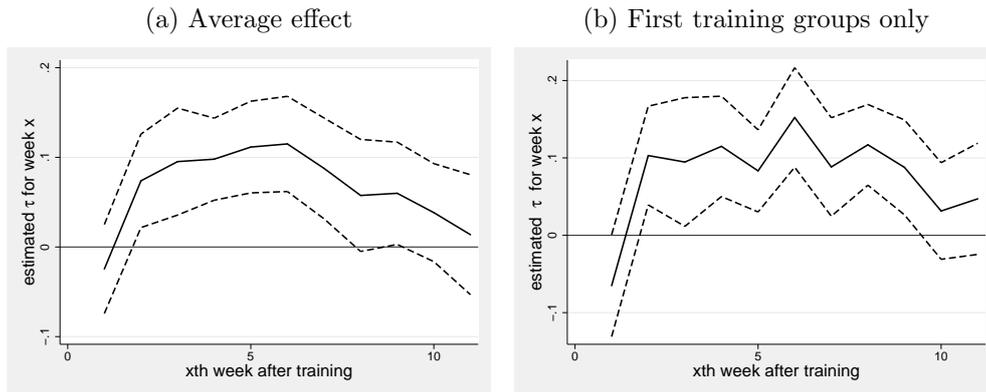
Note: This figure shows the estimated treatment effect on performance for each week after the training, controlling for a linear time trend. Panel (a) shows the estimated treatment effect from a regression including all agents. Panel (b) shows the treatment effect for first groups only. Week 0 denotes the training week. The dashed lines show the 95 percent confidence interval estimated from a regression of y_{it} on after-training week dummies.

Figure 4: Estimated payoff (in euros)
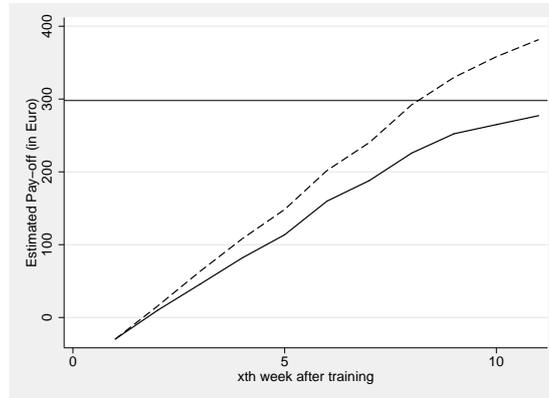
B   Tables

Table 1: Selection into the experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Selection into the experiment (N=179)				Initial assignment (N=86)			Estimation sample (N=74)		
Agents:	All	Ass.	Non-ass.	diff (2)-(3)	TG	CG	diff (5)-(6)	TG	CG	diff (8)-(9)
Gender	.2905	.2791	.3011	-.0220	.3513	.2245	.1269	.3824	.2750	.1074
(share of male agents)	(.4553)	(.4512)	(.4612)		(.4840)	(.4216)		(.4933)	(.4522)	
Age	32.595	35.575	29.431	6.143***	34.245	36.578	-2.332	34.070	36.146	-2.076
	(11.325)	(11.062)	(10.792)		(10.017)	(11.791)		(10.095)	(11.640)	
Tenure	2.646	4.312	1.106	3.206***	4.233	4.372	-.139	4.071	3.773	.298
(in years)	(3.610)	(4.019)	(2.310)		(3.796)	(4.218)		(3.810)	(3.984)	
Performance (y_{it})	.3307	.3716	.2928	.0789***	.3715	.3717	-.0002	.3596	.3582	.0013
	(.1108)	(.1017)	(.1058)		(.0884)	(.1116)		(.0779)	(.0893)	
Call quality (y_{it}^q)	.4541	.4676	.4416	-.0260	.4759	.4609	.0150	.4802	.4571	.0230
	(.1244)	(.1447)	(.1010)		(.1812)	(.1087)		(.1844)	(.1089)	
Composite productivity (y_{it}')	.2529	.2873	.2209	.0664***	.2846	.2895	.0049	.2759	.2780	-.0022
	(.0864)	(.0818)	(.0783)		(.0687)	(.0916)		(.0583)	(.0705)	
Average working hours	18.000	16.593	19.301	-2.708**	15.270	17.592	-2.322	15.706	17.350	-1.644
	(8.455)	(8.058)	(8.646)		(8.875)	(7.317)		(9.104)	(7.277)	
Share peak hours	.5541	.5304	.5773	-.0468*	.5445	.5197	.0248	.5553	.5286	.0266
	(.1766)	(.1946)	(.1547)		(.1972)	(.1939)		(.1926)	(.1970)	
Number of agents	179	86	93		37	49		34	40	

Difference significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations are shown in parentheses. Descriptive statistics are calculated before agents were assigned to the training (week 50/2008). TG=treatment group; CG=control group.

Table 2: Descriptive statistics of teams' first and second training groups (treatment group only)

Agents:	(1) First group	(2) Second group	(3) diff (2)-(3)
Gender	.3846	.3810	.0037
(share male agents)	(.5064)	(.4976)	
Age	35.659	33.085	2.575
	(10.759)	(9.799)	
Tenure	3.707	4.297	-.590
(in years)	(3.978)	(3.784)	
Performance (y_{it})	.3687	.3539	.0148
	(.0685)	(.0843)	
Call quality (y_{it}^q)	.4526	.4972	-.0446
	(.0798)	(.2269)	
Composite productivity (y_{it}')	.2844	.2706	.0139
	(.0538)	(.0616)	
Average working hours	18.539	13.952	4.586
	(10.674)	(7.743)	
Share peak hours	.5181	.5783	-.0602
	(.2242)	(.1720)	
Number of agents	13	21	

Difference significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations are shown in parentheses. Descriptive statistics are calculated before agents were assigned to the training (week 50/2008).

Table 3: Average performance post-treatment period (61 agents)

	(1) Treatment group	(2) Control group	(3) Difference TG-CG
Average performance (predicted)	.4148	.4027	.0122***
	(.0143)	(.0157)	
Number of agents	28	35	
Number of observations	296	409	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations in parentheses. Predicted performance is generated from a regression of performance on working hours of agent i in week t plus a linear time trend.

Table 4: Direct treatment effect

	(1)	(2)	(3)
Treatment dummy	.1127***	.1244***	.0882***
	(.0295)	(.0152)	(.0206)
Working hours	.0017	-.0012	-.0015
	(.0018)	(.0010)	(.0010)
Share peak hours	-.2564***	-.3486***	-.3416***
	(.0930)	(.0739)	(.0749)
Calls per FTE	.0002***	.0002**	.0001
	(.0001)	(.0001)	(.0001)
Time trend			.0023**
			(.0009)
Constant	-1.0008***	-.8770***	-1.0000***
	(.0782)	(.0582)	(.0888)
Individual fixed effects	No	Yes	Yes
Observations	1859	1859	1859
Number of agents	74	74	74
R^2	.0689	.6113	.6165

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level.

Table 5: Estimation of externalities

	(1)	(2)	(3)
Share of peers treated	.0513*		
	(.0275)		
Treatment dummy		.0992***	.0675**
		(.0213)	(.0287)
Constant	-1.1772***	-1.0548***	-.9924***
	(.0702)	(.1042)	(.1018)
Individual fixed effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Observations	1544	1352	1510
Number of agents	74	53	61
Number of agents (treatment group)	34	13	21
Number of agents (control group)	40	40	40
R^2	.5612	.6173	.5894

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level. *Share of peers treated* is calculated as the average share using information on the actual working hours in a specific week. The sample used for Column (1) is restricted to agents in an untreated state only (treatment dummy=0). Control variables include *working hours*, *share peak hours*, a linear time trend, and *calls per FTE*. The variable *share peak hours* is not used in Column (1).

Table 6: Pre-/post-differences in performance for other training programmes

	(1) short	(2) long
Length of courses		
Agents with short working hours	-.037	-.047***
	(.080)	(.012)
Number of observations	42	526
Agents with long working hours	-.006	-.040
	(.012)	(.034)
Number of observations	742	64

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The numbers show the differences in y_{it} between the first week after the training and the week before the training. The calculations are based on a larger data set from the same call centre. Short courses are defined as courses with less than 15 hours per week. Long courses are defined as courses with more than 25 hours per week. Short and long working hours are relative to the course length. Standard errors are in parentheses.

Table 7: Estimated treatment effect and estimation samples

Treatment dummy	.0773***
	(.0214)
Constant	-.9934***
	(.0993)
Individual fixed effects	Yes
Control variables	Yes
Observations	1532
Number of agents	55
Number of agents (treatment group)	26
Number of agents (control group)	29
R^2	.6178

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level. The samples is defined as all agents minus agents dropping out due to selection. Control variables include *working hours*, *share peak hours*, a linear time trend, and *calls per FTE*.

Table 8: The effect of training participation on call quality

Outcome	$\log(y_{it}^q)$	$\log(y_{it}')$	Grade 1	Grade 2	Grade 3
	(1)	(2)	(3)	(4)	(5)
Treatment dummy	.0425	.0985***	.8701**	1.5742***	-.4701
	(.0391)	(.0228)	(.3721)	(.4439)	(.9417)
Constant	-.5157***	-1.2132***	6.0604	3.2397	5.2156
	(.1374)	(.0876)	(3.8802)	(3.6075)	(5.4162)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes
Observations	1730	1730	112	115	107
Number of agents	74	74	14	15	15
R^2	.3384	.5763	.0283	.0474	.0405

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All standard errors are clustered at the agent level. Call quality y_{it}^q is measured as
1/(share of repeat calls*10). The composite productivity measure y_{it}' is defined as $y_{it}' = aht_{it} * (1-$ share of repeat calls).
Grades are given on 'knowledge of agent' (*Grade 1*), 'agent understood question' (*Grade 2*), and 'solution of the problem'
(*Grade 3*). Grades are given on a scale of 1 (very bad) to 10 (very good). Control variables include *working hours, share
peak hours*, a linear time trend, and *calls per FTE*.

Table 9: Direct treatment effect with interaction terms

Sample	Below-median tenure	Above-median tenure	All
	(1)	(2)	(3)
Treatment dummy	.0811***	.0977***	.1640***
	(.0243)	(.0325)	(.0378)
Working hours			-.0011
			(.0011)
Working hours × treatment			-.0037***
			(.0014)
Constant	-1.0476***	-1.0467***	-1.0027***
	(.0988)	(.1265)	(.0872)
Individual fixed effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Observations	908	951	1859
Number of agents	36	38	74
R^2	.1277	.1779	.1585

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. Column (1) and (2) show the baseline regression for
agents with below-median (1) and above-median (2) tenure, respectively. All standard errors are clustered at the agent
level. Control variables include *share peak hours*, a linear time trend, and *calls per FTE*.